

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2002 年 4 月 18 日 (18.04.2002)

PCT

(10) 国際公開番号  
WO 02/31810 A1

(51) 国際特許分類: G10L 13/00, 15/00, 15/22,  
B25J 5/00, 13/00, A63H 3/33, 11/00

Kazuo [JP/JP]. 広井 順 (HIROI, Jun) [JP/JP]. 小野  
木渡 (ONOGI, Wataru) [JP/JP]. 豊田 崇 (TOYODA,  
Takashi) [JP/JP]; 〒141-0001 東京都品川区北品川6丁  
目7番35号 ソニー株式会社内 Tokyo (JP).

(21) 国際出願番号: PCT/JP01/08898

(22) 国際出願日: 2001 年 10 月 10 日 (10.10.2001)

(74) 代理人: 稲本義雄 (INAMOTO, Yoshio); 〒160-0023 東  
京都新宿区西新宿7丁目5番8号 GOWA 西新宿ビル6  
階 Tokyo (JP).

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(81) 指定国 (国内): CN, KR, US.

(30) 優先権データ:  
特願 2000-310015  
2000 年 10 月 11 日 (11.10.2000) JP  
特願 2001-16872 2001 年 1 月 25 日 (25.01.2001) JP

(84) 指定国 (広域): ヨーロッパ特許 (AT, BE, CH, CY, DE,  
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

(71) 出願人 (米国を除く全ての指定国について): ソニー株  
式会社 (SONY CORPORATION) [JP/JP]; 〒141-0001  
東京都品川区北品川6丁目7番35号 Tokyo (JP).

添付公開書類:  
— 国際調査報告書

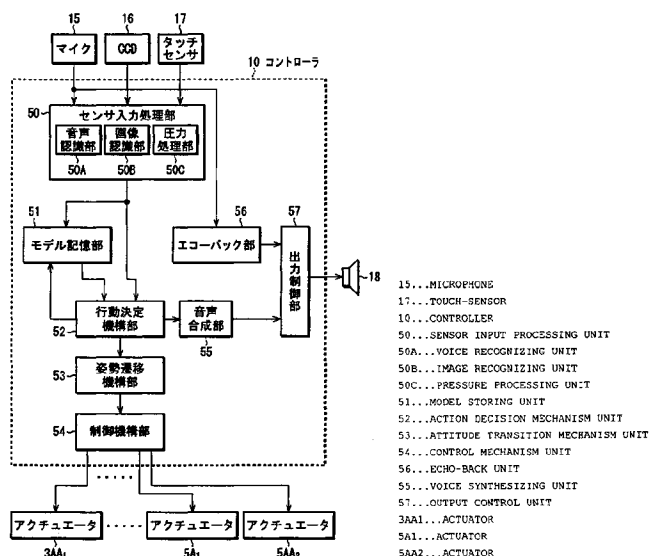
2 文字コード及び他の略語については、定期発行される  
各 PCT ガゼットの巻頭に掲載されている「コードと略語  
のガイダンスノート」を参照。

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 石井和夫 (ISHII,

(54) Title: ROBOT CONTROL APPARATUS

(54) 発明の名称: ロボット制御装置



(57) Abstract: A robot control apparatus. A user is allowed to recognize that his/her voice has been accepted. The user's voice inputted via a microphone (15) is voice-recognized by a voice recognizing unit (50A). An action deciding mechanism unit (52) decides the next robot action based on that voice recognition. Meanwhile, an echo-back unit (56) analyzes prosody information of the user's voice inputted via the microphone (15) and produces, on the basis of the analyzed prosody information, a voice whose prosody has been controlled, that is, an echo-back voice. This echo-back voice is applied through an output control unit (57) to a speaker (18) to be outputted thereby.



---

(57) 要約:

本発明は、ロボット制御装置に関する。音声を受け付けられていることを、ユーザに認識させる。マイク 15 に入力されたユーザの音声は、音声認識部 50 A で音声認識され、行動決定機構部 52 は、その音声認識結果に基づいて、ロボットの次の行動を決定する。一方、エコーバック部 56 は、マイク 15 に入力されたユーザの音声の韻律情報を分析し、その韻律情報に基づいて、韻律を制御した音声であるエコーバック音声を生成する。このエコーバック音声は、出力制御部 57 を介してスピーカ 18 に供給されて出力される。

## 明細書

## ロボット制御装置

## 技術分野

- 5      本発明は、ロボット制御装置に関し、特に、例えば、音声認識装置による音声認識結果に基づいて行動するロボットに用いて好適なロボット制御装置に関する。

## 背景技術

- 10      近年においては、例えば、玩具等として、ユーザが発した音声を音声認識し、その音声認識結果に基づいて、ある仕草をしたり、合成音を出力する等の行動を行うロボット（本明細書においては、ぬいぐるみ状のものを含む）が製品化されている。

- 15      しかしながら、ロボットにおいては、音声認識結果が得られても、何の行動も起こさない場合があり、この場合、ユーザは、ロボットにおいて音声認識されているのかどうか、あるいはロボットが故障しているのかどうか不安になる。

## 発明の開示

- 20      本発明は、このような状況に鑑みてなされたものであり、ロボットにおいて、ユーザの音声認識されていることを分かるようにすることで、ユーザの不安を解消すること等ができるようにするものである。

本発明のロボット制御装置は、音声認識されるユーザの音声の韻律情報を分析する分析手段と、韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成手段と、韻律制御音声データの出力を制御する出力制御手段とを備えることを特徴とする。

- 25      本発明のロボット制御方法は、音声認識されるユーザの音声の韻律情報を分析する分析ステップと、韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、韻律制御音声データの出力を制御

する出力制御ステップとを備えることを特徴とする。

- 本発明の記録媒体は、音声認識されるユーザの音声の韻律情報を分析する分析ステップと、韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、韻律制御音声データの出力を制御する出力
- 5 制御ステップとを備えるプログラムが記録されていることを特徴とする。

本発明のプログラムは、音声認識されるユーザの音声の韻律情報を分析する分析ステップと、韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、韻律制御音声データの出力を制御する出力制御ステップとを備えることを特徴とする。

- 10 本発明においては、音声認識されるユーザの音声の韻律情報が分析され、その韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データが生成されて出力される。

#### 図面の簡単な説明

- 15 図 1 は、本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

図 2 は、ロボットの内部構成例を示すブロック図である。

図 3 は、コントローラ 10 の機能的構成例を示すブロック図である。

図 4 は、音声認識部 50 A の構成例を示すブロック図である。

- 20 図 5 は、音声合成部 55 の構成例を示すブロック図である。

図 6 は、出力制御部 57 の処理を説明するフローチャートである。

図 7 は、エコーバック部 56 の構成例を示すブロック図である。

図 8 は、エコーバック部 56 によるエコーバック処理を説明するフローチャートである。

- 25 図 9 は、エコーバック部 56 による再エコーバック処理を説明するフローチャートである。

図 10 は、コントローラ 10 の他の機能的構成例を示すブロック図である。

発明を実施するための最良の形態

図 1 は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図 2 は、その電氣的構成例を示している。

- 5      本実施の形態では、ロボットは、例えば、犬等の四つ足の動物の形状のものとなっており、胴体部ユニット 2 の前後左右に、それぞれ脚部ユニット 3 A、3 B、3 C、3 D が連結されるとともに、胴体部ユニット 2 の前端部と後端部に、それぞれ頭部ユニット 4 と尻尾部ユニット 5 が連結されることにより構成されている。

- 10      尻尾部ユニット 5 は、胴体部ユニット 2 の上面に設けられたベース部 5 B から、2 自由度をもって湾曲または揺動自在に引き出されている。

胴体部ユニット 2 には、ロボット全体の制御を行うコントローラ 1 0、ロボットの動力源となるバッテリー 1 1、並びにバッテリーセンサ 1 2 および熱センサ 1 3 からなる内部センサ部 1 4 などが収納されている。

- 15      頭部ユニット 4 には、「耳」に相当するマイク（マイクロフォン）1 5、「目」に相当する CCD (Charge Coupled Device) カメラ 1 6、触覚に相当するタッチセンサ 1 7、「口」に相当するスピーカ 1 8 などが、それぞれ所定位置に配設されている。また、頭部ユニット 4 には、口の下顎に相当する下顎部 4 A が 1 自由度をもって可動に取り付けられており、この下顎部 4 A が動くことにより、ロボットの口の開閉動作が実現されるようになっている。

- 20      脚部ユニット 3 A 乃至 3 D それぞれの関節部分や、脚部ユニット 3 A 乃至 3 D それぞれと胴体部ユニット 2 の連結部分、頭部ユニット 4 と胴体部ユニット 2 の連結部分、頭部ユニット 4 と下顎部 4 A の連結部分、並びに尻尾部ユニット 5 と胴体部ユニット 2 の連結部分などには、図 2 に示すように、それぞれアクチュエータ 3 A A<sub>1</sub> 乃至 3 A A<sub>K</sub>、3 B A<sub>1</sub> 乃至 3 B A<sub>K</sub>、3 C A<sub>1</sub> 乃至 3 C A<sub>K</sub>、3 D A<sub>1</sub> 乃至 3 D A<sub>K</sub>、4 A<sub>1</sub> 乃至 4 A<sub>L</sub>、5 A<sub>1</sub> および 5 A<sub>2</sub> が配設されている。
- 25

頭部ユニット 4 におけるマイク 1 5 は、ユーザからの発話を含む周囲の音声（音）を集音し、得られた音声信号を、コントローラ 1 0 に送出する。CCD カ

メラ16は、周囲の状況を撮像し、得られた画像信号を、コントローラ10に送出する。

タッチセンサ17は、例えば、頭部ユニット4の上部に設けられており、ユーザからの「なでる」や「たたく」といった物理的な働きかけにより受けた圧力を  
5 検出し、その検出結果を圧力検出信号としてコントローラ10に送出する。

胴体部ユニット2におけるバッテリーセンサ12は、バッテリー11の残量を検出し、その検出結果を、バッテリー残量検出信号としてコントローラ10に送出する。  
熱センサ13は、ロボット内部の熱を検出し、その検出結果を、熱検出信号としてコントローラ10に送出する。

10 コントローラ10は、CPU(Central Processing Unit)10Aやメモリ10B等を内蔵しており、CPU10Aにおいて、メモリ10Bに記憶された制御プログラムが実行されることにより、各種の処理を行う。

即ち、コントローラ10は、マイク15や、CCDカメラ16、タッチセンサ17、バッテリーセンサ12、熱センサ13から与えられる音声信号、画像信号、  
15 圧力検出信号、バッテリー残量検出信号、熱検出信号に基づいて、周囲の状況や、ユーザからの指令、ユーザからの働きかけなどの有無を判断する。

さらに、コントローラ10は、この判断結果等に基づいて、続く行動を決定し、その決定結果に基づいて、アクチュエータ3A<sub>A1</sub>乃至3A<sub>A<sub>K</sub></sub>、3B<sub>A1</sub>乃至3B<sub>A<sub>K</sub></sub>、3C<sub>A1</sub>乃至3C<sub>A<sub>K</sub></sub>、3D<sub>A1</sub>乃至3D<sub>A<sub>K</sub></sub>、4A<sub>1</sub>乃至4A<sub>L</sub>、5A<sub>1</sub>、5A  
20 <sub>2</sub>のうちの必要なものを駆動させる。これにより、頭部ユニット4を上下左右に振らせたり、下顎部4Aを開閉させる。さらには、尻尾部ユニット5を動かさせたり、各脚部ユニット3A乃至3Dを駆動して、ロボットを歩行させるなどの行動を行わせる。

また、コントローラ10は、必要に応じて、合成音、あるいは後述するような  
25 エコーバック音声を生成し、スピーカ18に供給して出力させたり、ロボットの「目」の位置に設けられた図示しないLED(Light Emitting Diode)を点灯、消灯または点滅させる。

以上のようにして、ロボットは、周囲の状況等に基づいて自律的に行動をとるようになっている。

次に、図3は、図2のコントローラ10の機能的構成例を示している。なお、図3に示す機能的構成は、CPU10Aが、メモリ10Bに記憶された制御プログラムを実行することで実現されるようになっている。

コントローラ10は、特定の外部状態を認識するセンサ入力処理部50、センサ入力処理部50の認識結果を累積して、感情や、本能、成長の状態を表現するモデル記憶部51、センサ入力処理部50の認識結果等に基づいて、続く行動を決定する行動決定機構部52、行動決定機構部52の決定結果に基づいて、実際にロボットに行動を起こさせる姿勢遷移機構部53、各アクチュエータ3A<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>を駆動制御する制御機構部54、合成音を生成する音声合成部55、エコーバック音声を出力するエコーバック部56、並びに音声データの出力を制御する出力制御部57から構成されている。

センサ入力処理部50は、マイク15や、CCDカメラ16、タッチセンサ17等から与えられる音声信号、画像信号、圧力検出信号等に基づいて、特定の外部状態や、ユーザからの特定の働きかけ、ユーザからの指示等を認識し、その認識結果を表す状態認識情報を、モデル記憶部51および行動決定機構部52に通知する。

即ち、センサ入力処理部50は、音声認識部50Aを有しており、音声認識部50Aは、マイク15から与えられる音声信号について音声認識を行う。そして、音声認識部50Aは、その音声認識結果としての、例えば、「歩け」、「伏せ」、「ボールを追いかける」等の指令その他を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

また、センサ入力処理部50は、画像認識部50Bを有しており、画像認識部50Bは、CCDカメラ16から与えられる画像信号を用いて、画像認識処理を行う。そして、画像認識部50Bは、その処理の結果、例えば、「赤い丸いもの」や、「地面に対して垂直なかつ所定高さ以上の平面」等を検出したときには、

「ボールがある」や、「壁がある」等の画像認識結果を、状態認識情報として、モデル記憶部 5 1 および行動決定機構部 5 2 に通知する。

さらに、センサ入力処理部 5 0 は、圧力処理部 5 0 C を有しており、圧力処理部 5 0 C は、タッチセンサ 1 7 から与えられる圧力検出信号を処理する。そして、  
5 圧力処理部 5 0 C は、その処理の結果、所定の閾値以上で、かつ短時間の圧力を検出したときには、「たたかれた（しかられた）」と認識し、所定の閾値未満で、かつ長時間の圧力を検出したときには、「なでられた（ほめられた）」と認識して、その認識結果を、状態認識情報として、モデル記憶部 5 1 および行動決定機構部 5 2 に通知する。

10 モデル記憶部 5 1 は、ロボットの感情、本能、成長の状態を表現する感情モデル、本能モデル、成長モデルをそれぞれ記憶、管理している。

ここで、感情モデルは、例えば、「うれしさ」、「悲しさ」、「怒り」、「楽しさ」等の感情の状態（度合い）を、所定の範囲の値によってそれぞれ表し、センサ入力処理部 5 0 からの状態認識情報や時間経過等に基づいて、その値を変化  
15 させる。本能モデルは、例えば、「食欲」、「睡眠欲」、「運動欲」等の本能による欲求の状態（度合い）を、所定の範囲の値によってそれぞれ表し、センサ入力処理部 5 0 からの状態認識情報や時間経過等に基づいて、その値を変化させる。  
成長モデルは、例えば、「幼年期」、「青年期」、「熟年期」、「老年期」等の成長の状態（度合い）を、所定の範囲の値によってそれぞれ表し、センサ入力処  
20 理部 5 0 からの状態認識情報や時間経過等に基づいて、その値を変化させる。

モデル記憶部 5 1 は、上述のようにして感情モデル、本能モデル、成長モデルの値で表される感情、本能、成長の状態を、状態情報として、行動決定機構部 5 2 に送出する。

なお、モデル記憶部 5 1 には、センサ入力処理部 5 0 から状態認識情報が供給  
25 される他、行動決定機構部 5 2 から、ロボットの現在または過去の行動、具体的には、例えば、「長時間歩いた」などの行動の内容を示す行動情報が供給されるようになっており、モデル記憶部 5 1 は、同一の状態認識情報が与えられても、



行動情報が示すロボットの行動に応じて、異なる状態情報を生成するようになっている。

即ち、例えば、ロボットが、ユーザに挨拶をし、ユーザに頭を撫でられた場合には、ユーザに挨拶をしたという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部 51 に与えられ、この場合、モデル記憶部 51 では、「うれし  
5 しさ」を表す感情モデルの値が増加される。

一方、ロボットが、何らかの仕事を実行中に頭を撫でられた場合には、仕事を  
実行中であるという行動情報と、頭を撫でられたという状態認識情報とが、モデル  
記憶部 51 に与えられ、この場合、モデル記憶部 51 では、「うれしさ」を表  
10 す感情モデルの値は変化されない。

このように、モデル記憶部 51 は、状態認識情報だけでなく、現在または過去の  
ロボットの行動を示す行動情報も参照しながら、感情モデルの値を設定する。  
これにより、例えば、何らかのタスクを実行中に、ユーザが、いたずらするつもり  
で頭を撫でたときに、「うれしさ」を表す感情モデルの値を増加させるような、  
15 不自然な感情の変化が生じることを回避することができる。

なお、モデル記憶部 51 は、本能モデルおよび成長モデルについても、感情モデル  
における場合と同様に、状態認識情報および行動情報の両方に基づいて、その  
値を増減させるようになっている。また、モデル記憶部 51 は、感情モデル、  
本能モデル、成長モデルそれぞれの値を、他のモデルの値にも基づいて増減させ  
20 るようになっている。

行動決定機構部 52 は、センサ入力処理部 50 からの状態認識情報や、モデル  
記憶部 51 からの状態情報、時間経過等に基づいて、次の行動を決定し、決定さ  
れた行動の内容を、行動指令情報として、姿勢遷移機構部 53 に送出する。

即ち、行動決定機構部 52 は、ロボットがとり得る行動をステート（状態）(s  
25 tate)に対応させた有限オートマトンを、ロボットの行動を規定する行動モデル  
として管理しており、この行動モデルとしての有限オートマトンにおけるステ  
ートを、センサ入力処理部 50 からの状態認識情報や、モデル記憶部 51 における

感情モデル、本能モデル、または成長モデルの値、時間経過等に基づいて遷移させ、遷移後のステートに対応する行動を、次にとるべき行動として決定する。

ここで、行動決定機構部 5 2 は、所定のトリガ(trigger)があったことを検出すると、ステートを遷移させる。即ち、行動決定機構部 5 2 は、例えば、現在の  
5 ステートに対応する行動を実行している時間が所定時間に達したときや、特定の状態認識情報を受信したとき、モデル記憶部 5 1 から供給される状態情報が示す感情や、本能、成長の状態の値が所定の閾値以下または以上になったとき等に、ステートを遷移させる。

なお、行動決定機構部 5 2 は、上述したように、センサ入力処理部 5 0 からの  
10 状態認識情報だけでなく、モデル記憶部 5 1 における感情モデルや、本能モデル、成長モデルの値等にも基づいて、行動モデルにおけるステートを遷移させることから、同一の状態認識情報が入力されても、感情モデルや、本能モデル、成長モデルの値（状態情報）によっては、ステートの遷移先は異なるものとなる。

その結果、行動決定機構部 5 2 は、例えば、状態情報が、「怒っていない」こと  
15 と、および「お腹がすいていない」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「お手」という行動をとらせる行動指令情報を生成し、これを、姿勢遷移機構部 5 3 に送出する。

また、行動決定機構部 5 2 は、例えば、状態情報が、「怒っていない」こと、  
20 および「お腹がすいている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「手のひらをぺろぺろなめる」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部 5 3 に送出する。

また、行動決定機構部 5 2 は、例えば、状態情報が、「怒っている」ことを表  
25 している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、状態情報が、「お腹がすいている」ことを表しているても、また、「お腹がすいていない」ことを表しているても、「ぶいと横を向く」

ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部 5 3 に送出する。

5       なお、行動決定機構部 5 2 には、モデル記憶部 5 1 から供給される状態情報が示す感情や、本能、成長の状態に基づいて、遷移先のステートに対応する行動のパラメータとしての、例えば、歩行の速度や、手足を動かす際の動きの大きさおよび速度などを決定させることができ、この場合、それらのパラメータを含む行動指令情報が、姿勢遷移機構部 5 3 に送出される。

10       また、行動決定機構部 5 2 では、上述したように、ロボットの頭部や手足等を動作させる行動指令情報の他、ロボットに発話を行わせる行動指令情報も生成される。ロボットに発話を行わせる行動指令情報は、音声合成部 5 5 に供給されるようになり、音声合成部 5 5 に供給される行動指令情報には、音声合成部 5 5 に生成させる合成音に対応するテキスト等が含まれる。そして、音声合成部 5 5 は、行動決定部 5 2 から行動指令情報を受信すると、その行動指令情報に含まれるテキストに基づき、合成音を生成し、出力制御部 5 7 を介して、スピーカ 1 8 に供給して出力させる。これにより、スピーカ 1 8 からは、例えば、ロボットの鳴き声、さらには、「お腹がすいた」等のユーザへの各種の要求、「何？」等のユーザの呼びかけに対する応答その他の音声出力が行われる。

20       姿勢遷移機構部 5 3 は、行動決定機構部 5 2 から供給される行動指令情報に基づいて、ロボットの姿勢を、現在の姿勢から次の姿勢に遷移させるための姿勢遷移情報を生成し、これを制御機構部 5 4 に送出する。

      ここで、現在の姿勢から次に遷移可能な姿勢は、例えば、胴体や手や足の形状、重さ、各部の結合状態のようなロボットの物理的形状と、関節が曲がる方向や角度のようなアクチュエータ 3 A<sub>1</sub>乃至 5 A<sub>1</sub> および 5 A<sub>2</sub> の機構とによって決定される。

25       また、次の姿勢としては、現在の姿勢から直接遷移可能な姿勢と、直接には遷移できない姿勢とがある。例えば、4 本足のロボットは、手足を大きく投げ出して寝転んでいる状態から、伏せた状態へ直接遷移することはできるが、立った状

態へ直接遷移することはできず、一旦、手足を胴体近くに引き寄せて伏せた姿勢になり、それから立ち上がるという２段階の動作が必要である。また、安全に実行できない姿勢も存在する。例えば、４本足のロボットは、その４本足で立っている姿勢から、両前足を挙げてバンザイをしようとする、簡単に転倒してしまう。

このため、姿勢遷移機構部５３は、直接遷移可能な姿勢をあらかじめ登録しておき、行動決定機構部５２から供給される行動指令情報が、直接遷移可能な姿勢を示す場合には、その行動指令情報を、そのまま姿勢遷移情報として、制御機構部５４に送出する。一方、行動指令情報が、直接遷移不可能な姿勢を示す場合には、姿勢遷移機構部５３は、遷移可能な他の姿勢に一旦遷移した後に、目的の姿勢まで遷移させるような姿勢遷移情報を生成し、制御機構部５４に送出する。これによりロボットが、遷移不可能な姿勢を無理に実行しようとする事態や、転倒するような事態を回避することができるようになっている。

制御機構部５４は、姿勢遷移機構部５３からの姿勢遷移情報にしたがって、アクチュエータ３ＡＡ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>を駆動するための制御信号を生成し、これを、アクチュエータ３ＡＡ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>に送出する。これにより、アクチュエータ３ＡＡ<sub>１</sub>乃至５Ａ<sub>１</sub>および５Ａ<sub>２</sub>は、制御信号にしたがって駆動し、ロボットは、自律的に行動を起こす。

エコーバック部５６は、マイク１５から与えられ、音声認識部５０Ａで音声認識される音声信号を監視しており、その音声信号を復唱するような音声（以下、適宜、エコーバック音声という）を生成して出力する。このエコーバック音声は、出力制御部５７を介して、スピーカ１８に供給されて出力される。

出力制御部５７には、音声合成部５５からの合成音のデジタルデータと、エコーバック部５６からのエコーバック音声のデジタルデータとが供給されるようになっており、それらのデジタルデータを、アナログの音声信号にＤ／Ａ変換し、スピーカ１８に供給して出力させる。また、出力制御部５７は、音声合成部５５からの合成音と、エコーバック部５６からのエコーバック音声の、スピー

カ 1 8 への出力が競合した場合に、その競合を調整する。即ち、エコーバック部 5 6 からのエコーバック音声の出力は、行動決定機構部 5 2 の制御にしたがって音声合成部 5 5 が行う合成音の出力とは独立に行われるようになっており、このため、エコーバック音声の出力と合成音の出力とは競合する場合がある。そこで、  
5 出力制御部 5 7 は、その競合の調停を行う。

次に、図 4 は、図 3 の音声認識部 5 0 A の構成例を示している。

マイク 1 5 からの音声信号は、A/D (Analog Digital) 変換部 2 1 に供給される。A/D 変換部 2 1 では、マイク 1 5 からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データに A/D 変換される。この  
10 音声データは、特徴抽出部 2 2 および音声区間検出部 2 7 に供給される。

特徴抽出部 2 2 は、そこに入力される音声データについて、適当なフレームごとに、例えば、M F C C (Mel Frequency Cepstrum Coefficient) 分析を行い、その分析の結果得られる M F C C を、特徴パラメータ（特徴ベクトル）として、マッチング部 2 3 に出力する。なお、特徴抽出部 2 2 では、その他、例えば、線形  
15 予測係数、ケプストラム係数、線スペクトル対、所定の周波数帯域ごとのパワー（フィルタバンクの出力）等を、特徴パラメータとして抽出することが可能である。

マッチング部 2 3 は、特徴抽出部 2 2 からの特徴パラメータを用いて、音響モデル記憶部 2 4、辞書記憶部 2 5、および文法記憶部 2 6 を必要に応じて参照しながら、マイク 1 5 に入力された音声（入力音声）を、例えば、連続分布 HMM  
20 (Hidden Markov Model) 法に基づいて音声認識する。

即ち、音響モデル記憶部 2 4 は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここでは、連続分布 HMM 法に基づいて音声認識を行うので、音響モデルとしては、HMM (Hidden Markov Model) が用いられる。辞書記憶部 2 5 は、認識対象の各単語について、  
25 その発音に関する情報（音韻情報）が記述された単語辞書を記憶している。文法記憶部 2 6 は、辞書記憶部 2 5 の単語辞書に登録されている各単語が、どのよう

に連鎖する（つながる）かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法（C F G）や、統計的な単語連鎖確率（N-gram）などに基づく規則を用いることができる。

マッチング部 2 3 は、辞書記憶部 2 5 の単語辞書を参照することにより、音響  
5 モデル記憶部 2 4 に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部 2 3 は、幾つかの単語モデルを、文法記憶部 2 6 に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、連続分布HMM法によって、マイク 1 5 に入力された音声を認識する。即ち、マッチング部 2 3 は、特徴抽出部 2 2 が出力する時系列の特徴パラメータが観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の音韻情報（読み）を、音声の認識結果として出力する。

より具体的には、マッチング部 2 3 は、接続された単語モデルに対応する単語列について、各特徴パラメータの出現確率を累積し、その累積値をスコアとして、  
15 そのスコアを最も高くする単語列の音韻情報を、音声認識結果として出力する。

以上のようにして出力される、マイク 1 5 に入力された音声の認識結果は、状態認識情報として、モデル記憶部 5 1 および行動決定機構部 5 2 に出力される。

なお、音声区間検出部 2 7 は、AD変換部 2 1 からの音声データについて、特徴抽出部 2 2 がMFCC分析を行うのと同様のフレームごとに、例えば、パワー  
20 を算出している。さらに、音声区間検出部 2 7 は、各フレームのパワーを、所定の閾値と比較し、その閾値以上のパワーを有するフレームで構成される区間を、ユーザの音声が入力されている音声区間として検出する。そして、音声区間検出部 2 7 は、検出した音声区間を、特徴抽出部 2 2 とマッチング部 2 3 に供給しており、特徴抽出部 2 2 とマッチング部 2 3 は、音声区間のみを対象に処理を行う。  
25 次に、図 5 は、図 3 の音声合成部 5 5 の構成例を示している。

テキスト生成部 3 1 には、行動決定機構部 5 2 が出力する、音声合成の対象とするテキストを含む行動指令情報が供給されるようになっており、テキスト生成

部 3 1 は、辞書記憶部 3 4 や生成用文法記憶部 3 5 を参照しながら、その行動指令情報に含まれるテキストを解析する。

即ち、辞書記憶部 3 4 には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、生成用文法記憶部 3 5 には、辞  
5 書記憶部 3 4 の単語辞書に記述された単語について、単語連鎖に関する制約等の生成用文法規則が記憶されている。そして、テキスト生成部 3 1 は、この単語辞書および生成用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部 3 2 で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の  
10 韻律情報や、各単語の発音等の音韻情報などがある。

テキスト生成部 3 1 で得られた情報は、規則合成部 3 2 に供給され、規則合成部 3 2 では、音素片記憶部 3 6 を参照しながら、テキスト生成部 3 1 に入力されたテキストに対応する合成音の音声データ（デジタルデータ）が生成される。

即ち、音素片記憶部 3 6 には、例えば、C V (Consonant, Vowel) や、V C V、C V C 等の形で音素片データが記憶されており、規則合成部 3 2 は、テキスト生成部 3 1 からの情報に基づいて、必要な音素片データを接続し、さらに、音素片データの波形を加工することによって、ポーズ、アクセント、イントネーション等を適切に付加し、これにより、テキスト生成部 3 1 に入力されたテキストに対応  
20 する合成音の音声データを生成する。

以上のようにして生成された音声データは、出力制御部 5 7（図 3）を介して、スピーカ 1 8 に供給され、これにより、スピーカ 1 8 からは、テキスト生成部 3 1 に入力されたテキストに対応する合成音が出力される。

なお、図 3 の行動決定機構部 5 2 では、上述したように、行動モデルに基づいて、次の行動が決定されるが、合成音として出力するテキストの内容は、ロボッ  
25 トの行動と対応付けておくことが可能である。

即ち、例えば、ロボットが、座った状態から、立った状態になる行動には、テ

キスト「よっこいしょ」などを対応付けておくことが可能である。この場合、ロボットが、座っている姿勢から、立つ姿勢に移行するときに、その姿勢の移行に同期して、合成音「よっこいしょ」を出力することが可能となる。

次に、図6のフローチャートを参照して、図3の出力制御部57の処理（出力  
5 制御処理）について説明する。

出力制御部57は、音声合成部55からの合成音を受信すると、エコーバック部56からエコーバック音声を受信していない限り、音声合成部55からの合成音を、スピーカ18に供給して出力させる。

そして、出力制御部57は、エコーバック部56からエコーバック音声を受信  
10 した場合、図6のフローチャートにしたがった出力制御処理を行う。

即ち、出力制御部57は、エコーバック部56からエコーバック音声を受信した場合、まず最初に、ステップS1において、音声合成部55から合成音が出力されているかどうかを判定し、出力されていないと判定した場合、ステップS2に進み、エコーバック部56から受信したエコーバック音声を、スピーカ18に  
15 供給して出力させ、処理を終了する。

一方、ステップS1において、音声合成部55から合成音が出力されていると判定された場合、即ち、音声合成部55からの合成音の出力と、エコーバック部56からのエコーバック音声の出力とが競合した場合、ステップS3に進み、出力制御部57は、エコーバック音声の出力を制限し、音声合成部55からの合成  
20 音を、スピーカ18に供給して、出力させ、処理を終了する。

なお、出力制御部57はエコーバック音声の出力中に、音声合成部55から合成音を受信した場合、エコーバック音声の出力を中止し、音声合成部55からの合成音の出力を優先させる。

以上のように、出力制御部57は、音声合成部55からの合成音の出力と、エ  
25 コーバック部56からのエコーバック音声の出力とが競合した場合、音声合成部55からの合成音の出力を優先させる。これは、音声合成部55からの合成音の出力が、行動決定機構部52からの行動指令情報に基づくものであり、ロボット



の行動であるのに対して、エコーバック音声の出力は、ユーザの発話を復唱するものであることから、音声合成部 55 からの合成音の方が、エコーバック音声に比較して、ユーザに対する重要度が高いと考えられるためである。

但し、出力制御部 57 では、例えば、エコーバック音声の出力を、音声合成部 55 からの合成音の出力より優先させることも可能であるし、また、例えば、エコーバック音声と、音声合成部 55 からの合成音のうち、先に受信した方の出力を優先させるようにすること等も可能である。

次に、図 7 は、図 3 のエコーバック部 56 の構成例を示している。

マイク 15 からの音声信号は、AD変換部 41 に供給される。AD変換部 41 では、マイク 15 からのアナログ信号である音声信号がサンプリング、量子化され、デジタル信号である音声データに A/D 変換される。この音声データは、韻律分析部 42 および音声区間検出部 46 に供給される。

韻律分析部 42 は、そこに入力される音声データを、適当なフレームごとに音響分析することにより、例えば、ピッチ周波数やパワー等といった音声データの韻律情報を抽出する。この韻律情報は、音生成部 43 に供給される。

即ち、韻律分析部 42 は、例えば、人間の一般的なピッチ周期よりも長い 32 ms 等を 1 フレームとして、AD変換部 41 からの音声データのピッチ周波数とパワーを、フレーム単位で求める。

具体的には、韻律分析部 42 は、第  $j$  フレームのパワー  $P(j)$  を、次式にしたがって求める。

$$P(j) = \sum x_i^2 \quad \dots (1)$$

但し、式 (1) において、 $x_i$  は、第  $j$  フレームの先頭から  $i$  番目の音声データのサンプル値を表し、 $\sum$  は、第  $j$  フレームの音声データのサンプルすべてについてのサメーションを表す。

また、韻律分析部 42 は、第  $j$  フレームの音声データのサンプルについて、自己相関値を演算し、その自己相関値を最大にする、0 でないずれ量 (時間)  $\tau$  を

求めて、次式により、第  $j$  フレームのピッチ周波数  $f(j)$  を演算する。

$$f(j) = 1/\tau$$

・・・ (2)

音生成部 4 3 は、韻律分析部 4 2 からの韻律情報に基づいて、韻律を制御した  
5 エコーバック音声を生成する。

即ち、音生成部 4 3 は、例えば、韻律分析部 4 2 からの韻律情報と同一の韻律  
を有する、音韻のない音声（以下、適宜、無音韻音声という）を、サイン(sin)  
波を重ねることにより生成し、エコーバック音声として、出力部 4 4 に供給す  
る。

10 なお、韻律情報としての、例えば、ピッチ周波数とパワーから音声データを生  
成する方法については、例えば、鈴木、石井、竹内、「非分節音による反響的な  
模倣とその心理的影響」、情報処理学会論文誌、vol. 41, No. 5, pp1328-1337, May,  
2000 や、特開 2000-181896 号公報等に、その一例が記載されている。

出力部 4 4 は、音生成部 4 3 からのエコーバック音声のデータを、メモリ 4 5  
15 に記憶させるとともに、出力制御部 5 7（図 3）に出力する。

音声区間検出部 4 6 は、AD変換部 4 1 からの音声データについて、図 4 の音  
声区間検出部 2 7 における場合と同様の処理を行うことにより、音声区間を検出  
し、韻律分析部 4 2 と音生成部 4 3 に供給する。これにより、韻律分析部 4 2 と  
音生成部 4 3 では、音声区間のみを対象に処理が行われる。

20 なお、図 7 の AD変換部 4 1 または音声区間検出部 4 6 と、図 4 の AD変換部  
2 1 または音声区間検出部 2 7 とは、それぞれ兼用することが可能である。

以上のように構成されるエコーバック部 5 6 では、例えば、図 8 のフローチャ  
ートにしたがったエコーバック処理が行われる。

即ち、まず最初に、ステップ S 1 1 において、音声区間検出部 4 6 が、AD変  
25 換部 4 1 の出力に基づいて、音声区間であるかどうかを判定し、音声区間でない  
と判定した場合、処理を終了し、再度、ステップ S 1 1 からのエコーバック処理  
を再開する。

また、ステップS 1 1において、音声区間であると判定された場合、即ち、マイク 1 5に、ユーザの音声が入力された場合、ステップS 1 2に進み、韻律分析部 4 2は、AD変換部 4 1の出力、即ち、マイク 1 5に入力されたユーザの音声を音響分析することにより、その韻律情報を取得し、音生成部 4 3に供給する。

- 5 音生成部 4 3は、ステップS 1 3において、韻律分析部 4 2からの韻律情報と同一の韻律を有する無音韻音声を生成し、エコーバック音声として、出力部 4 4に供給する。

- 出力部 4 4は、ステップS 1 4において、音生成部 4 3からのエコーバック音声のデータを、メモリ 4 5に記憶させ、ステップS 1 5に進み、そのエコーバック  
10 ク音声を、出力制御部 5 7（図 3）に出力して、処理を終了する。

これにより、エコーバック音声は、出力制御部 5 7を介して、スピーカ 1 8に供給されて出力される。

従って、この場合、スピーカ 1 8からは、ユーザが発した音声から、その音韻を無くしたものが、エコーバック音声として出力される。

- 15 このエコーバック音声は、音声認識部 5 0 Aにおいて音声認識の対象とされるユーザの音声を復唱するようなものであり、このようなエコーバック音声が出力される結果、ユーザは、エコーバック音声を聴くことにより、ロボットにおいて、自身の音声を受け付けられたことを認識することができる。従って、ロボットが、ユーザからの音声に対する応答として、何の行動も起こさない場合（音声認識部  
20 5 0 Aにおいて、ユーザの音声が正しく認識されている場合と、誤って認識されている場合の両方を含む）であっても、ユーザにおいて、ロボットが故障しているといったような勘違いをすること等を防止することができる。

- さらに、エコーバック音声は、ユーザが発した音声そのものではなく、その音声の音韻をなくしたものであるため、ユーザには、ロボットが、ユーザの音声を  
25 理解し、自身の声で復唱しているかのように聞こえる。また、エコーバック音声は、ユーザが発した音声と同一の韻律を有することから、その韻律から得られる感情を表現するようなものとなっており、その結果、ユーザには、ロボットが、

ユーザの感情を理解しているかのように聞こえる。従って、ロボットにおいて、ユーザの音声を、単に録音して再生しているのではなく、理解しているかのような印象を、ユーザに与えることができる。

なお、ここでは、音生成部 4 3 において、サイン波を重畳することによって、  
5 エコーバック音声を生成するようにしたが、その他、例えば、エコーバック音声は、ロボットの鳴き声となるような複雑な波形を用意しておき、その波形をつなぎ合わせることによって生成することが可能である。さらに、エコーバック音声としては、例えば、ユーザの音声を構成する音素を認識し、その音素列によって構成される音韻を有するようなものを生成することが可能である。また、エコー  
10 バック音声は、例えば、ユーザの音声について、ケプストラム係数を得て、そのケプストラム係数をタップ係数とするディジタルフィルタによって生成すること等も可能である。

但し、エコーバック音声は、ユーザの音声に似過ぎると、ロボットにおいて、ユーザの音声を、単に録音して再生しているかのような、いわば興ざめた印象  
15 を、ユーザに与えかねないので、エコーバック音声は、ユーザの音声に、あまり似たものにしない方が望ましい。

また、上述の場合には、音生成部 4 3 において、ユーザの音声の韻律と同一の韻律を有するエコーバック音声を生成するようにしたが、音生成部 4 3 には、ユーザの音声の韻律に加工を加えた韻律を有するエコーバック音声を生成させるこ  
20 とも可能である。

即ち、例えば、上述したようにして、韻律分析部 4 2 において、第  $j$  フレームのパワー  $P(j)$  とピッチ周波数  $f(j)$  が求められる場合には、音生成部 4 3 では、そのパワー  $P(j)$  とピッチ周波数  $f(j)$  を用い、例えば、次式にしたがって、エコーバック音声  $y(t)$  を生成することができる。

$$25 \quad y(t) = \log(P(j)) \sin(2\pi N f(j)t) \quad \dots (3)$$

但し、式 (3) において、 $N$  は、正の整数である。

式（３）によれば、第  $j$  フレームの音声に対して、そのパワー  $P(j)$  を対数によって非線形にし、かつ、ピッチ周波数  $f(j)$  を  $N$  倍にしたエコーバック音声  $y(t)$  が生成される。

5   なお、経験的には、 $N$  を、例えば 6 等とした場合に、鳴き声らしく、かつ聞きやすいエコーバック音声を得られる。

また、上述の場合においては、ユーザの音声の音声区間と同一の時間長のエコーバック音声が出力されることとなるが、エコーバック音声は、ユーザの音声の音声区間と同一の時間長ではなく、異なる時間長とすることが可能である。

10   即ち、例えば、式（３）にしたがって、エコーバック音声を生成する場合においては、パワー  $P(j)$  やピッチ周波数  $f(j)$  を補間または間引きすること等によって、ユーザの音声の音声区間よりも長いまたは短い時間長のエコーバック音声を生成するようにすることが可能である。

15   なお、経験的には、エコーバック音声は、ユーザの音声の音声区間の、例えば、 $3/2$  倍等の、長めの時間長にする方が、ユーザに自然な感じを与えるものとなる。

ところで、図 7 のエコーバック部 56 においては、メモリ 45 に、エコーバック音声記憶されるようになっており、出力部 44 は、このメモリ 45 に記憶されたエコーバック音声を、必要に応じて読み出して出力する再エコーバック処理を行う。

20   即ち、再エコーバック処理では、図 9 のフローチャートに示すように、出力部 44 は、ステップ S 21 において、所定のイベントが生じたかどうかを判定し、生じていないと判定した場合、ステップ S 21 に戻る。

25   また、ステップ S 21 において、所定のイベントが生じたと判定された場合、ステップ S 22 に進み、出力部 44 は、メモリ 45 に記憶された所定のエコーバック音声のデータを読み出し、出力制御部 57 を介して、スピーカ 18 に供給して、ステップ S 21 に戻る。

従って、再エコーバック処理によれば、所定のイベントが生じた場合に、メモ

り45に記憶された所定のエコーバック音声、スピーカ18から出力される。

ここで、所定のイベントとしては、例えば、ある値を設定しておき、乱数を発生させて、その乱数が、設定した値に一致した場合や、外部からの入力がない状態が、所定の時間連続した場合等を採用することができる。この場合、ユーザからすれば、ロボットに対して、何の刺激も与えていないのに、ロボットが、突然喋りだしたように見えるので、ロボットのエンタテインメント性を向上させることができる。

また、メモリ45には、エコーバック音声のデータを、行動決定機構部52が出力する所定の行動指令情報と対応付けて記憶させておき、行動決定機構部52が行動指令情報を出力した場合に、その行動指令情報に対応付けられているエコーバック音声を、メモリ45から読み出すようにすることも可能である。この場合、行動決定機構部52が出力する行動指令情報に対応する行動に同期して、その行動指令情報に対応付けられているエコーバック音声出力されることになる。

次に、図10は、コントローラ10の他の構成例を示している。なお、図中、図3における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図10のコントローラ10は、エコーバック部56において、モデル記憶部51に記憶されたモデルが参照可能になっている他は、図3における場合と同様に構成されている。

コントローラ10が図10に示したように構成される場合には、エコーバック部56の音生成部43において、図7に点線で示すように、モデル記憶部51に記憶された感情モデルや、本能モデル、成長モデルが参照される。

そして、音生成部43は、感情モデルや、本能モデル、成長モデルによって表されるロボットの感情や、本能、成長の状態に基づいて、ユーザの音声から抽出した韻律を制御したエコーバック音声を生成する。

即ち、音生成部43は、例えば、ロボットの喜びの度合い、または悲しみの度合いが大きい場合には、その度合いに応じて、ピッチ周波数が高い、または低いエコーバック音声をそれぞれ生成する。また、音生成部43は、例えば、ロボッ

トの怒りの度合いが大きい場合には、その度合いに応じて、発話時間の短いエコーバック音声を生成する。その他、音生成部 4 3 には、例えば、成長の度合いが高い場合には、語尾のピッチ周波数を高くして、質問を行うようなエコーバック音声を生成させること等が可能であり、この場合、ユーザに、次の発話を促すことができる。

ここで、図 9 における場合には、再エコーバック処理において、乱数や外部からの入力等を、所定のイベントとして、メモリ 4 5 に記憶された所定のエコーバック音声を再生するようにしたが、図 1 0 の実施の形態のように、エコーバック部 5 6 の音生成部 4 3 において、モデル記憶部 5 1 に記憶された感情モデルや、本能モデル、成長モデルを参照する場合には、音生成部 4 3 で生成されたエコーバック音声を、そのエコーバック音声が生じられたときの感情モデル等の値と対応付けて、メモリ 4 5 に記憶しておき、モデル記憶部 5 1 に記憶された感情モデル等の値が、メモリ 4 5 に記憶されたものと一致したことを、所定のイベントとして、その一致した感情モデル等の値と対応付けられているエコーバック音声を、メモリ 4 5 から再生して出力するようにすることが可能である。

この場合、再エコーバック処理によれば、メモリ 4 5 に記憶されたエコーバック音声は、感情モデル等の値が、そのエコーバック音声が生じられたときのものと同一になったときに再生されることとなるから、ロボットは、ある感情等の状態になると、同一のエコーバック音声を出力することとなる。そして、その結果、ユーザは、そのようなエコーバック音声を聴くことによって、ロボットの感情等の状態を把握することが可能となるので、ユーザとロボットとの間で、いわば、感情等を共有することが可能となる。

なお、行動決定機構部 5 2 と出力制御部 5 7 とは、それらの間で通信を行うように構成することができ、この場合、行動決定機構部 5 2 には、出力制御部 5 7 からスピーカ 1 8 に音声データが出力されるときに、下顎部 4 A (図 1) を動かすアクチュエータを駆動する行動指令情報を生成させるようにすることができる。この場合、スピーカ 1 8 からの音声出力に同期して、下顎部 4 A が動くので、ユ

一々に、ロボットが実際に喋っているかのような印象を与えることができる。

さらに、行動決定機構部 5 2 では、下顎部 4 A の動きを、出力制御部 5 7 からスピーカ 1 8 に出力される音声データの、例えば、パワー等の韻律情報に基づいて制御することが可能である。即ち、例えば、音声データのパワーが大の場合には、下顎部 4 A の動きを大きくし、これにより、ロボットの口の開き具合を大きくする一方、音声データのパワーが小の場合には、下顎部 4 A の動きを小さくし、これにより、ロボットの口の開き具合を小さくすることが可能である。この場合、スピーカ 1 8 から出力される音の大きさに対応して、ロボットの口の開き具合が変化するので、ロボットの喋り方に、よりリアリティをもたせることができる。

10    以上、本発明を、エンターテイメント用のロボット（疑似ペットとしてのロボット）に適用した場合について説明したが、本発明は、これに限らず、例えば、産業用のロボット等の各種のロボットに広く適用することが可能である。また、本発明は、現実世界のロボットだけでなく、例えば、液晶ディスプレイ等の表示装置に表示される仮想的なロボットにも適用可能である。

15    さらに、本実施の形態においては、上述した一連の処理を、CPU 1 0 A にプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

20    なお、プログラムは、あらかじめメモリ 1 0 B（図 2）に記憶させておく他、フロッピーディスク、CD-ROM (Compact Disc Read Only Memory)、MO (Magneto optical) ディスク、DVD (Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納（記録）しておくことができる。そして、このようなりムーバブル記録媒体を、いわゆるパッケージソフトウェアとして提供し、ロボット（メモリ 1 0 B）にインストールするように行うことができる。

25    また、プログラムは、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、無線で転送したり、LAN (Local Area Network)、インターネットといったネットワークを介して、有線で転送し、メモリ 1 0 B にインストールす



ることができる。

この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、メモリ 10 B に、容易にインストールすることができる。

- 5      ここで、本明細書において、CPU 10 A に各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

- 10     また、プログラムは、1 の CPU により処理されるものであっても良いし、複数の CPU によって分散処理されるものであっても良い。

#### 産業上の利用可能性

- 本発明によれば、音声認識されるユーザの音声の韻律情報が分析され、その韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データが生成
- 15     されて出力される。従って、ユーザは、韻律制御音声データを聴くことにより、ロボットにおいて、自身の音声を受け付けられたことを認識することができる。

## 請求の範囲

1. 少なくとも音声認識結果に基づいて行動するロボットを制御するロボット制御装置であって、  
音声認識されるユーザの音声の韻律情報を分析する分析手段と、  
5 前記韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成手段と、  
前記韻律制御音声データの出力を制御する出力制御手段と  
を備えることを特徴とするロボット制御装置。
2. 前記生成手段は、音韻のない前記韻律制御音声データを生成する  
10 ことを特徴とする請求の範囲第1項に記載のロボット制御装置。
3. 前記ロボットの感情、本能、または成長の状態を表すモデルを記憶するモデル記憶手段をさらに備え、  
前記生成手段は、前記モデル記憶手段に記憶されたモデルが表す前記ロボットの感情、本能、または成長の状態にも基づいて、前記韻律制御音声データを生成  
15 する  
ことを特徴とする請求の範囲第1項に記載のロボット制御装置。
4. 前記韻律制御音声データを記憶するデータ記憶手段をさらに備え、  
前記出力制御手段は、所定のイベントが生じたときに、前記データ記憶手段に記憶された韻律制御音声データを出力する  
20 ことを特徴とする請求の範囲第1項に記載のロボット制御装置。
5. 前記ロボットの行動を制御する行動制御手段をさらに備え、  
前記出力制御手段は、前記行動制御手段による制御によって、音声の出力が行われる場合に、前記韻律制御音声データの出力を制限する  
ことを特徴とする請求の範囲第1項に記載のロボット制御装置。
- 25 6. 前記韻律制御音声データの出力に同期して、前記ロボットに、その口部を動かす行動をとらせる行動制御手段をさらに備える  
ことを特徴とする請求の範囲第1項に記載のロボット制御装置。

7. 前記行動制御手段は、前記韻律制御音声データに基づいて、前記ロボットの口部の動き具合を制御する

ことを特徴とする請求の範囲第6項に記載のロボット制御装置。

8. 少なくとも音声認識結果に基づいて行動するロボットを制御するロボット  
5 制御方法であって、

音声認識されるユーザの音声の韻律情報を分析する分析ステップと、

前記韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、

前記韻律制御音声データの出力を制御する出力制御ステップと

10 を備えることを特徴とするロボット制御方法。

9. 少なくとも音声認識結果に基づいて行動するロボットを制御するロボット制御処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

音声認識されるユーザの音声の韻律情報を分析する分析ステップと、

15 前記韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、

前記韻律制御音声データの出力を制御する出力制御ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

20 10. 少なくとも音声認識結果に基づいて行動するロボットを制御するロボット制御処理を、コンピュータに行わせるプログラムであって、

音声認識されるユーザの音声の韻律情報を分析する分析ステップと、

前記韻律情報に基づいて、韻律を制御した音声データである韻律制御音声データを生成する生成ステップと、

25 前記韻律制御音声データの出力を制御する出力制御ステップと

を備えることを特徴とするプログラム。

1/9

図 1

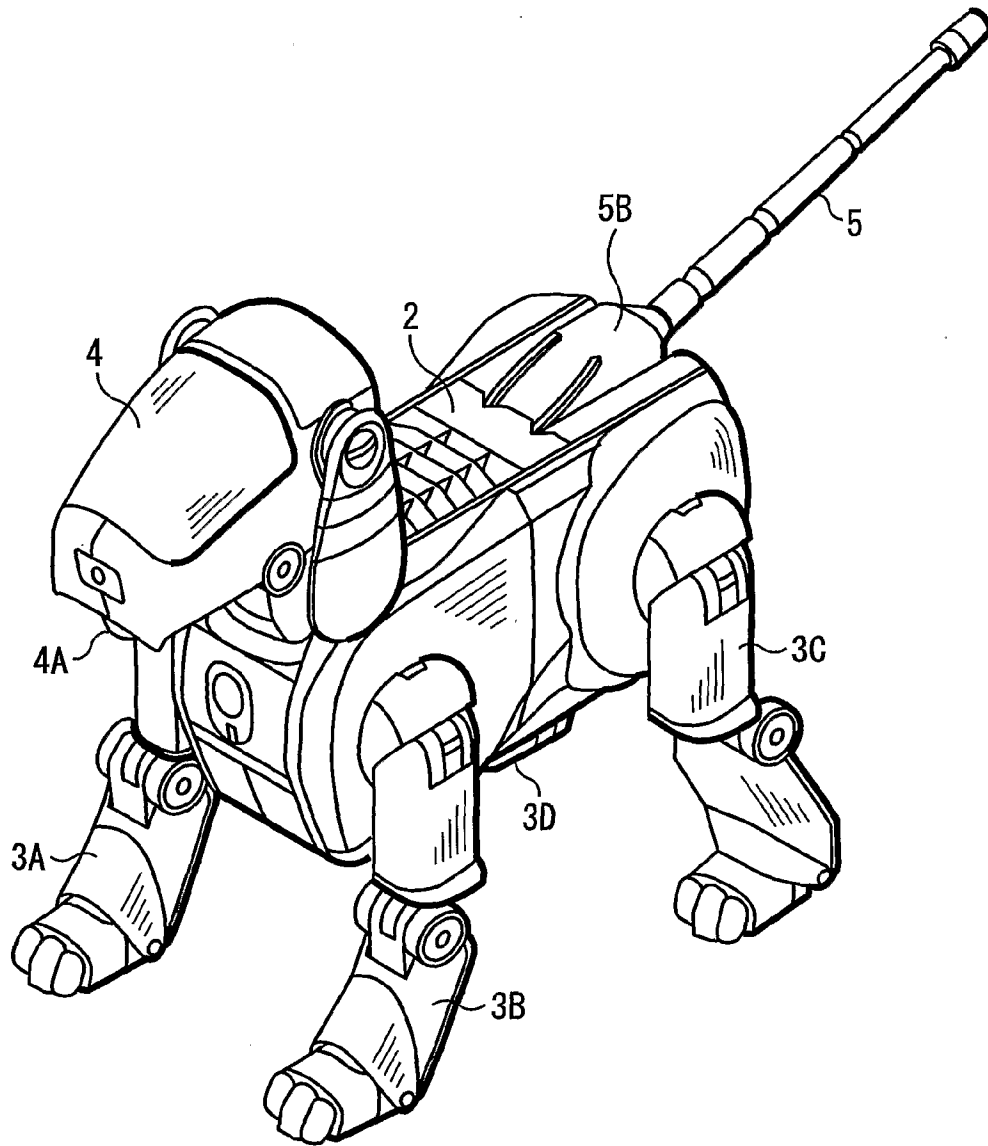
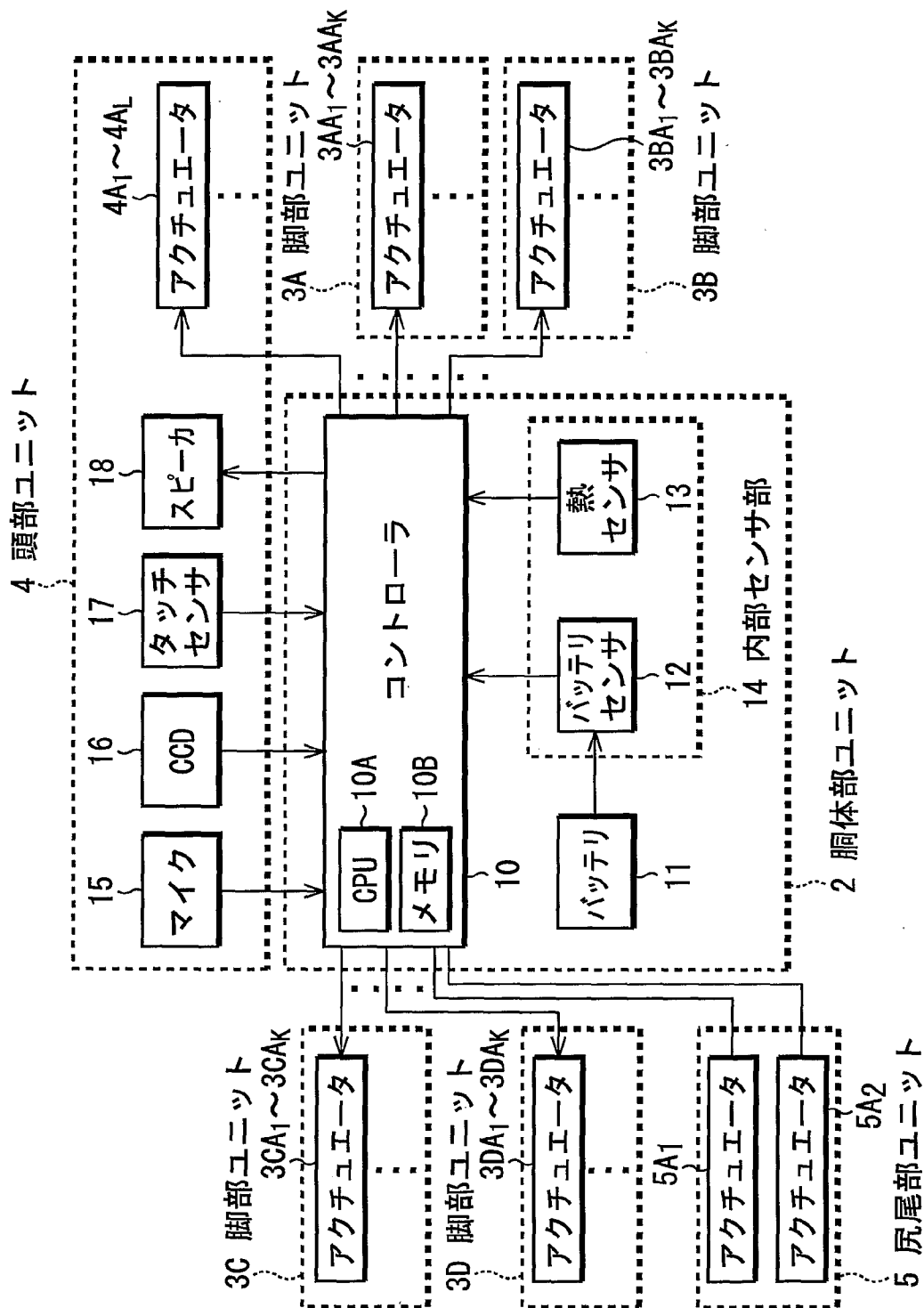
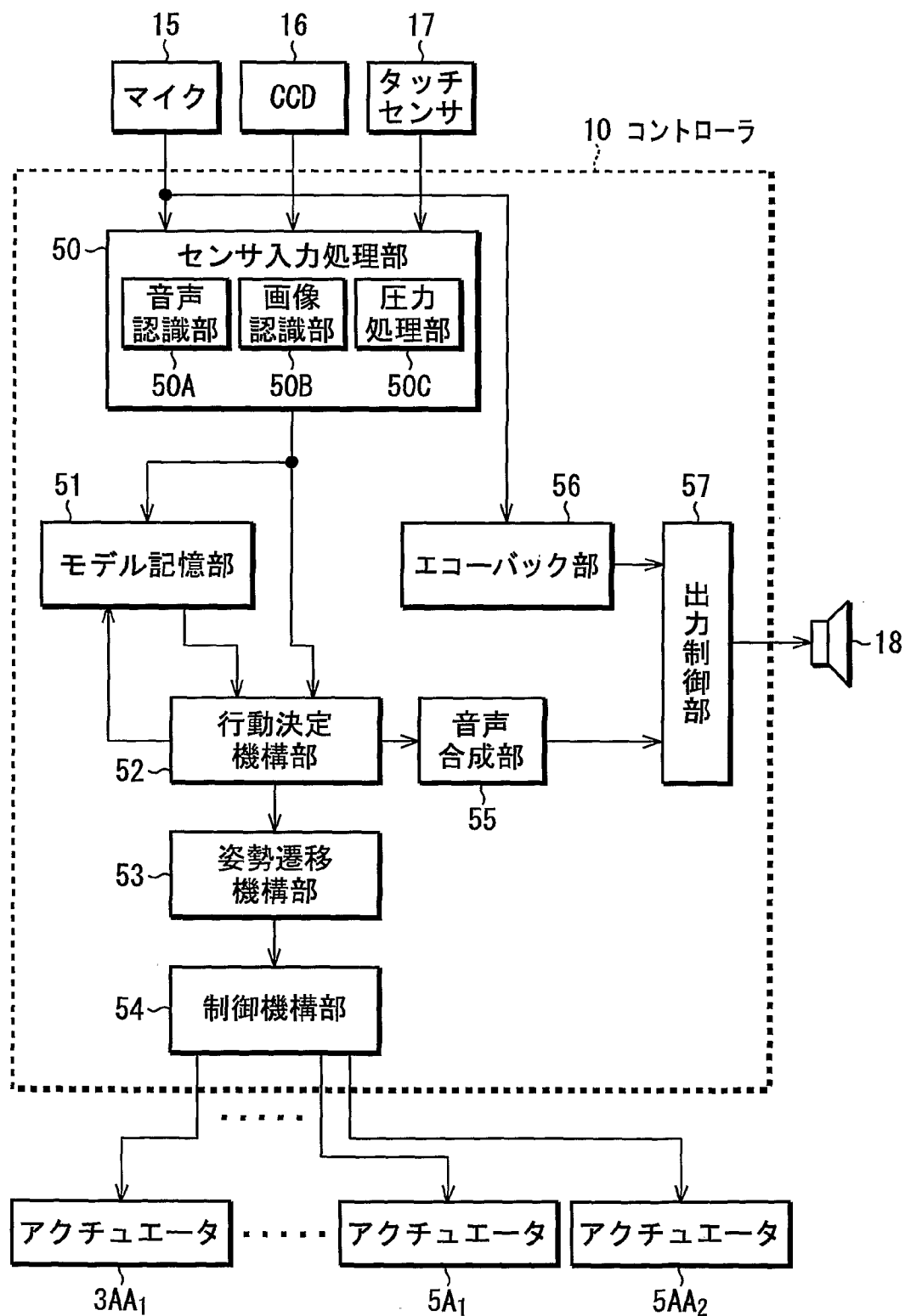


図 2



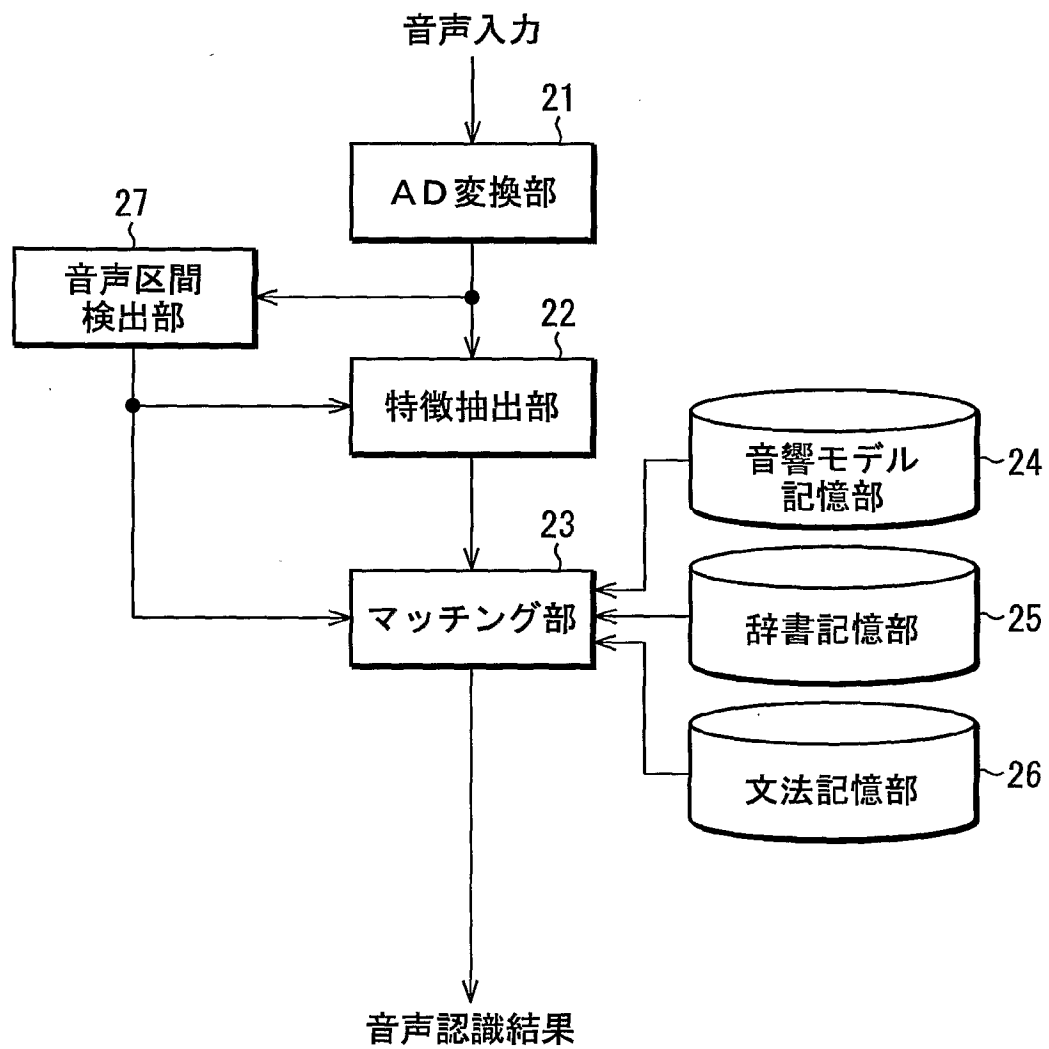
3/9

図 3



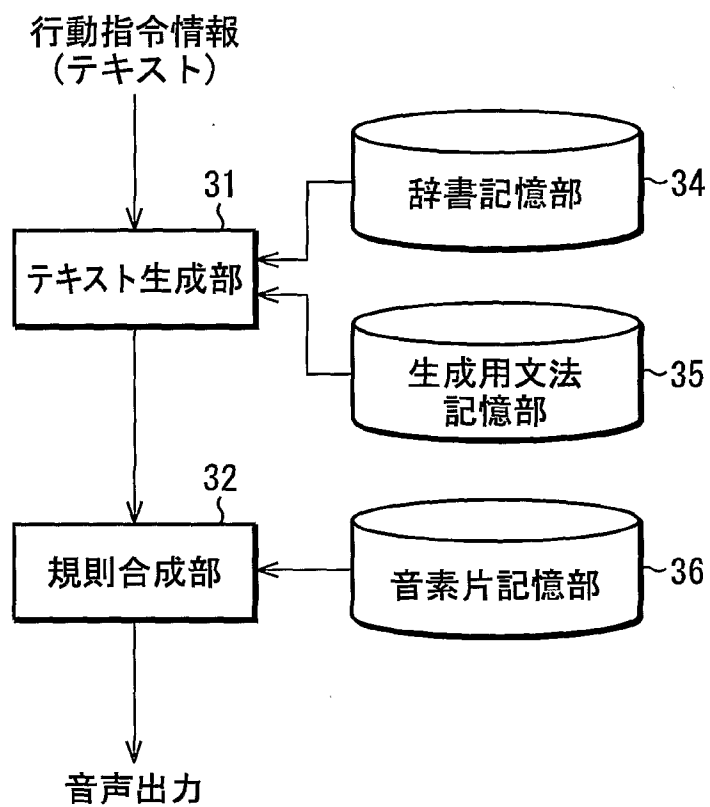
4/9

図 4



5/9

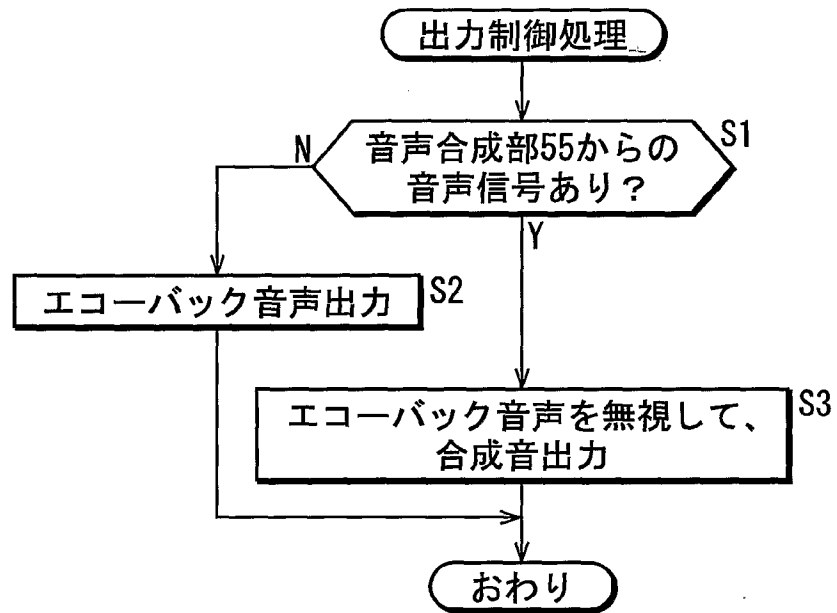
図 5





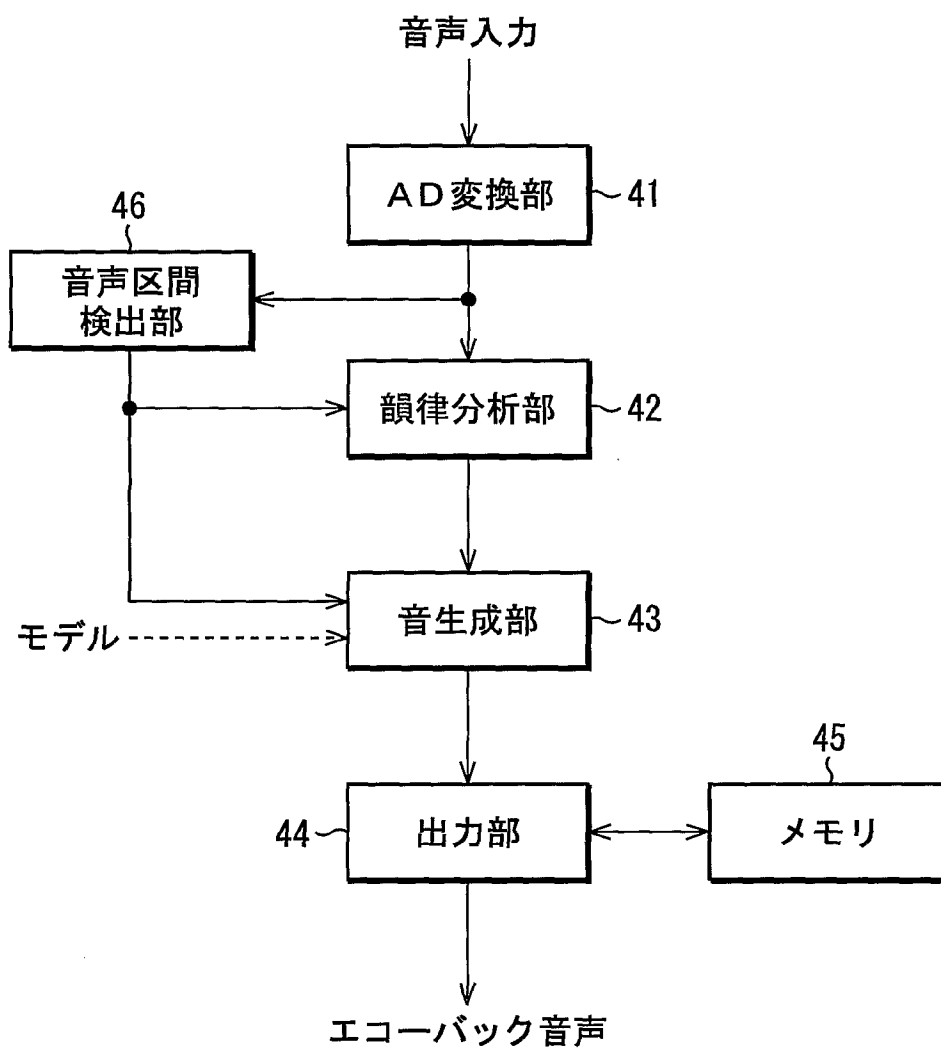
6/9

図 6



7/9

図 7



8/9

図 8

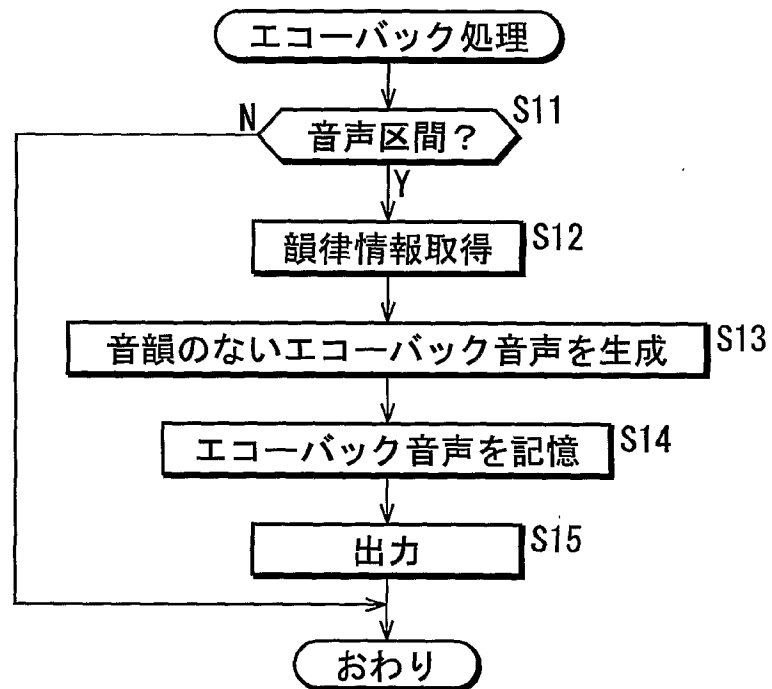
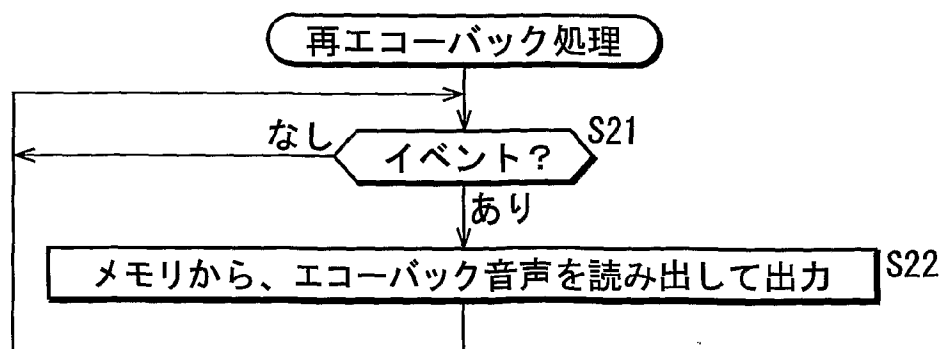
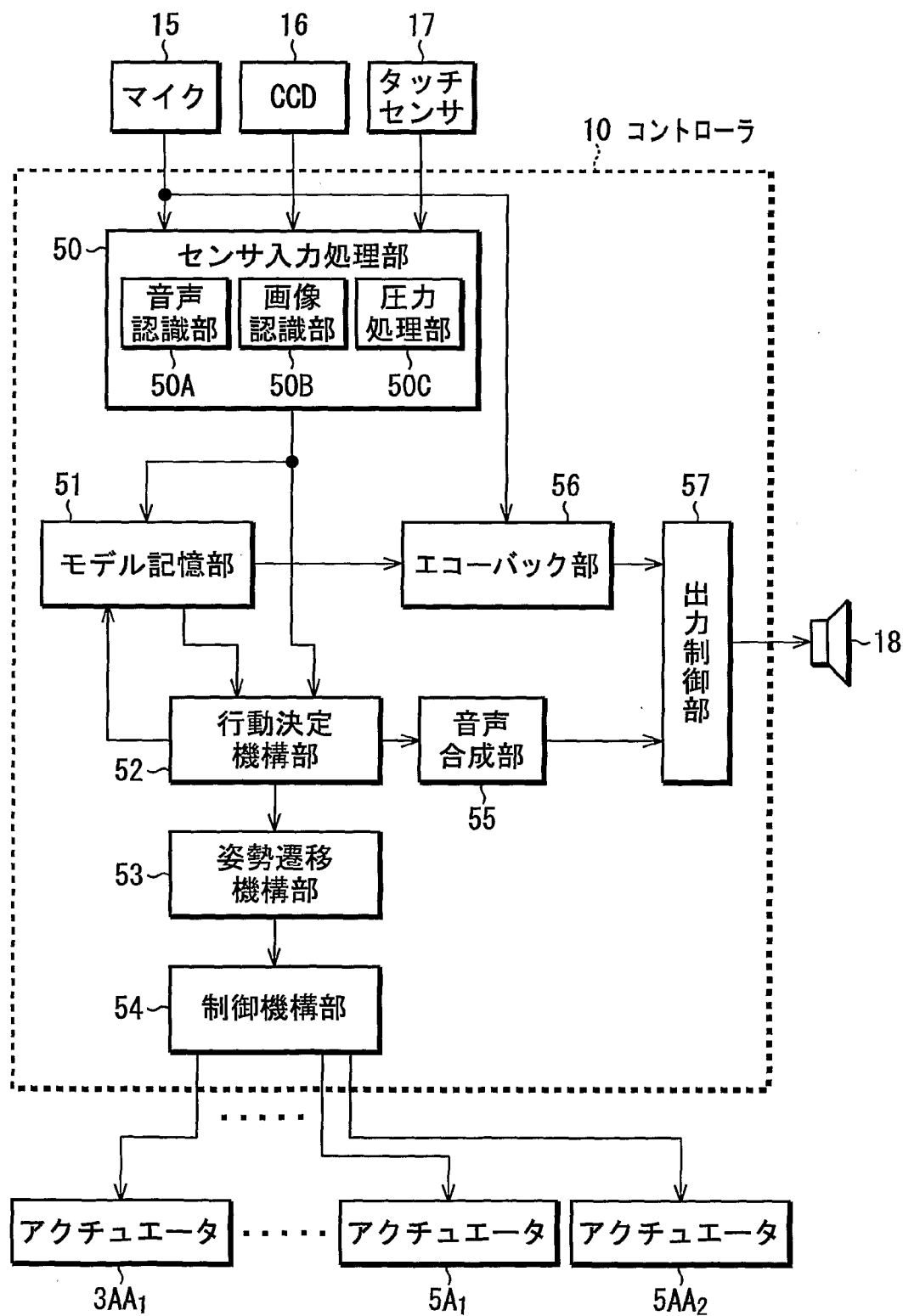


図 9



9/9

図10



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP01/08898

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> G10L13/00, 15/00, 15/22, B25J5/00, 13/00, A63H3/33, 11/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G10L13/00, 15/00, 15/22, B25J5/00, 13/00, A63H3/33, 11/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1926-1995 Toroku Jitsuyo Shinan Koho 1994-2001  
 Kokai Jitsuyo Shinan Koho 1971-2001 Jitsuyo Shinan Toroku Koho 1996-2001

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2000-135186 A (Kabushiki Kaisha YM Creation), 16 May, 2000 (16.05.00), Full text; all drawings (Family: none)	1,3-5,8-10
Y	JP 2000-259601 A (Masami KATO), 22 September, 2000 (22.09.00), Full text; all drawings (Family: none)	1,3-7,8-10
Y	JP 9-152884 A (Fujitsu Ten, Limited), 10 June, 1997 (10.06.97), Full text; all drawings (Family: none)	1,3-7,8-10
Y	JP 10-111786 A (Sharp Corporation), 28 April, 1998 (28.04.98), Full text; all drawings (Family: none)	1,3-7,8-10
Y	EP 0730261 A2 (Seiko Epson Corporation), 04 September, 1996 (04.09.96), Full text; all drawings & US 5802488 A & JP 8-297498 A	3

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 02 November, 2001 (02.11.01)	Date of mailing of the international search report 13 November, 2001 (13.11.01)
Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP01/08898

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 9-62480 A (Casio Computer Co., Ltd.), 07 March, 1997 (07.03.97), Full text; all drawings (Family: none)	3
A	JP 2000-254359 A (Casio Computer Co., Ltd.), 19 September, 2000 (19.09.00), Full text; all drawings (Family: none)	1
A	JP 2000-29492 A (Hitachi, Ltd.), 28 January, 2000 (28.01.00), Full text; all drawings (Family: none)	1
A	JP 4-37997 B2 (Toshiba Corporation), 23 June, 1992 (23.06.92), Full text; all drawings (Family: none)	1
A	Joho Shori Gakkai Ronbunshi, Noriko SUZUKI et al., "Hi-Bunsetsu-on ni yoru Hankyou-teki na Mohou to sono Sihnri-teki Eikyou", Vol.41, No.5, pages 1328 to 1338, (May, 2000)	1-10

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G10L13/00, 15/00, 15/22, B25J5/00, 13/00  
A63H3/33, 11/00

## B. 調査を行った分野

## 調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G10L13/00, 15/00, 15/22, B25J5/00, 13/00  
A63H3/33, 11/00

## 最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年  
日本国公開実用新案公報 1971~2001年  
日本国登録実用新案公報 1994~2001年  
日本国実用新案登録公報 1996~2001年

## 国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P 2000-135186 A (株式会社ワイ・エム・クリエ イション) 16.5月.2000 (16.05.00) 全文, 全図 (ファミリーな し)	1, 3-5, 8-10
Y	J P 2000-259601 A (加藤誠巳) 22.9月.2000 (22. 09.00) 全文, 全図 (ファミリーなし)	1, 3-7, 8-10
Y	J P 9-152884 A (富士通テン株式会社) 10.6月.1997 (10.06.97) 全文, 全図 (ファミリーなし)	1, 3-7, 8-10

☒ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」 口頭による開示、使用、展示等に言及する文献  
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

## の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
「&」 同一パテントファミリー文献

国際調査を完了した日

02.11.01

国際調査報告の発送日

13.11.01

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)  
郵便番号100-8915  
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

榎本 剛

5C

9379

電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P 10-111786 A (シャープ株式会社) 28.4月.1998 (28.04.98) 全文, 全図 (ファミリーなし)	1,3-7,8-10
Y	E P 0730261 A2 (Seiko Epson Corporation) 04.9月. 1996 (04.09.96) 全文, 全図 & US 5802488 A & J P 8-297498 A	3
Y	J P 9-62480 A (カシオ計算機株式会社) 7.3月.1997 (07.03.97) 全文, 全図 (ファミリーなし)	3
A	J P 2000-254359 A (カシオ計算機株式会社) 19. 9月.2000 (19.09.00) 全文, 全図 (ファミリーなし)	1
A	J P 2000-29492 A (株式会社日立製作所) 28.1月. 2000 (28.01.00) 全文, 全図 (ファミリーなし)	1
A	J P 4-37997 B2 (株式会社東芝) 23.6月.1992 (23. 06.92) 全文, 全図 (ファミリーなし)	1
A	情報処理学会論文誌, 鈴木紀子他「非分節音による反響的な模倣と その心理的影響」 Vol.41, No.5, p.1328-1338 (5月.2000年)	1-10